Transportation Mobility Factor Extraction Using Image Recognition Techniques

Pittipol Kantavat
Department of Computer Engineering,
Faculty of Engineering,
Chulalongkorn University
Bangkok, Thailand
pittipol.k@chula.ac.th

Boonserm Kijsirikul
Department of Computer Engineering,
Faculty of Engineering,
Chulalongkorn University
Bangkok, Thailand
boonserm.k@chula.ac.th

Yuji Iwahori
Department of Computer Science,
College of Engineering,
Chubu University
Kasugai, Japan
iwahori@isc.chubu.ac.jp

Yoshitsugu Hayashi
Center for Sustainable Development
and Global Smart City,
Chubu University
Kasugai, Japan
y-hayashi@isc.chubu.ac.jp

Teerapong Panboonyuen
Department of Computer Engineering,
Faculty of Engineering,
Chulalongkorn University
Bangkok, Thailand
teerapong.panboonyuen@gmail.com

Peerapon Vateekul
Department of Computer Engineering,
Faculty of Engineering,
Chulalongkorn University
Bangkok, Thailand
peerapon.v@chula.ac.th

Witsarut Achariyaviriya
Department of Constructional
Engineering, Graduate School of
Engineering, Chubu University
Kasugai, Japan
witsarut.ac@gmail.com

Abstract—**For an urban development, the Quality of Life (QOL) of people in the city is a vital issue that should be considered. There are many researches in QOL topics that use questionnaire survey approach. These studies yield very useful information for city development planning. As the Artificial Intelligence technologies are developed very fast recently, they are applied to solve many transportation problems. In this paper, we propose a method that automatically extract mobility indicators using two image recognition techniques: Semantic Segmentation and Object Recognition. Because the mobility is an important factor in QOL evaluation, our work can be used to enhance a performance and reduce a data gathering cost of the QOL evaluation.**

Keywords—Quality of Life (QOL), Transportation Mobility, Image Recognition, Semantic Segmentation, Object Detection

I. Introduction

In many countries, big cities are developing very fast. This leads to increasing in the number of population and many consequences to be considered, including the quality of life (QOL) of people in the cities.

Several research topics of QOL have been conducted. The main issues are what are important factors influencing QOL and how to evaluate QOL using the related factors. In many studies, transportation is an important factor that directly impacts QOL. Schneider [1] studied people behaviors in Minnesota, USA and reported that transportation was important because it connected people to their destinations. Doi et al. [2] proposed a QOL evaluation method based on accessibility and social interaction. Nakamura et al. [3] evaluated megacities development by combining a land-use transportation model with QOL index using Bangkok as a case study. Nakamura et al. [4] analyzed difference of QOL in station areas by comparing Nagoya to Bangkok and suggested the importance of developing efficient transportation systems. Gu et al. [5] investigated a methodology to evaluate QOL of people in Nanging and considered a convenience in transportation network accessing as an evaluating factor. These studies in QOL not only provide us urban people's insights, but also give useful information for urban development. However, the data gathering process was costly and time-consuming. Moreover, the gathering information could be changed over time.

In recent years, the image recognition technology has been developed rapidly. It has been used for solving problems in various practical fields including transportation. Chowdhury et al. [6] proposed a model for traffic management by counting the number of vehicles on the road. Hua and Anastasiu [7] designed a tracking algorithm for a smart traffic network running in a real-time environment. Putri et al. [8] estimated traffic density using a video processing model. Kim et al. [9] proposed a framework for vehicle tracking using the Faster R-CNN algorithm. Osman et al. [10] designed an intelligent traffic management system for road junctions using image processing. Chen and Huang [11] modeled a moving vehicle tracking system based on traffic surveillance systems. Ali et al. [12] proposed an intelligent and autonomous traffic management system for reducing traffic congestion problems. Munajat et al. [13] designed a road condition detection algorithm using RGB histogram filtering. Dinh et al. [14] constructed a traffic jam warning system based on coarse data in Vietnam. Trivedi et al. [15] designed a vehicle counting module for smart traffic management in a small city. Delavarian and Maarouzi [16] proposed a multi-object tracking system for intersection using multilayer image sequences. Mithun et al. [17] developed a video-based intelligent traffic management system that can track objects and also their flow directions. These systems overcome the traditional traffic management system because they work more efficient, while human operation cost is lower. In addition, we can collect data log and analyze people behavior in order to plan traffic management policy in the future.

In this paper, we apply two types of image processing techniques to extract useful information from a video dataset. The first one is *Semantic Segmentation* to recognize areas of observed objects. The second one is *Object Recognition* which can detect and count the number of observed objects. The acquired information can be used to figure transportation mobility, which is a vital factor affecting QOL of people. Using an automated algorithm for data gathering process will reduce survey costs. Moreover, we can gather new up-to-date data as much as we need to evaluate QOL in the future.

This paper is organized as follows. Section II discusses overview of QOL and transportation mobility. Section III reviews image recognition techniques that are used in this paper. Section IV proposes our method. Section V provides experimental details and results. Section VI discusses about possible future works. Section VII summarizes our research.

II. TRANSPORTATION AND QUALITY OF LIFE

A. Accessibility and Quality of Life

In context of QOL evaluation, accessibility is employed in many papers. In [2], the authors concluded that accessibility were directly and indirectly related to five QOL elements: safety and security, economic opportunity, service and cultural opportunity, spatial amenity and environmental benignity. Accessibility was mentioned again in [4] as one of four elements that were included in the questionnaire: accessibility, amenity, safety and cost. The research also found that peoples in both Bangkok and Nagoya valued accessibility as the most important factor for them. Using facility data points from BaiduMap and road/subway networks from OpenStreetMap, QOL index was calculated based on distribution of accessible values in education service, shopping services and medical care service [5]. According to these papers, we can see that accessibility is a vital issue for QOL evaluation.

B. Accessibility vs Mobility

As shown in the previous section, accessibility is a vital factor in QOL evaluating process. Mobility is another factor affecting QOL. Both accessibility and mobility are related to travelling quality. Accessibility means the quality in travelling. It focuses on travel time, travel cost, and travel options. On the other hand, mobility is the ability and level of ease of moving goods and services. It focuses on traffic congestion, obstacles, and the number of lanes. Hence, accessibility and mobility are the same thing in different point of view.

In our research, as we will apply image recognition techniques, we will measure our experimental results in the view of mobility. To illustrate more, we will detect and count the number of observed objects from one destination to another; but we will not measure travel time or cost consumed in travelling between two destinations.

III. IMAGE RECOGNITION

Among researches in Artificial Intelligence, Image Recognition is growing very fast, especially after emerging of the Convolutional neural networks (CNNs), a Deep Learning for Image Recognition, have become the dominant machine learning approach for semantic segmentation and object detection.

In this paper, two computer vision types are used: (i) semantic segmentation (ii) object detection. The difference between these two types is that semantic segmentation classifies each pixel into a set of categories; whereas object detection localizes and classifies an object using a bounding box.

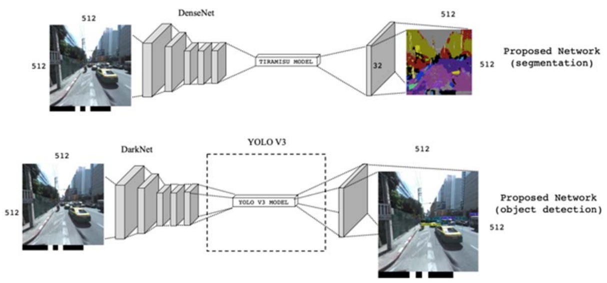


Fig. 1. An overview of our deep learning architectures

A. Semantic Segmentation

Semantic segmentation classifies all the pixels of an image into meaningful classes of objects and helps determine the relations between objects, as well as the context of objects in the image.

DenseNet [18] is originally used for image classification. Several advantages of DenseNet are: (i) it is more efficient in the parameter usage (ii) it performs deep supervision thanks to short paths to all feature maps in the deep learning architecture, and (iii) all layers can easily access their preceding layers making it easy to reuse the information from previously computed feature maps.

In this paper, we present Fully Convolutional DenseNet for semantic segmentation. Because of the segmentation task, Fully Convolutional DenseNet (also known as The One Hundred Layers Tiramisu, in short as Tiramisu) [19] has up-sampling layers on the contrary to the DenseNet. Tiramisu uses a downsampling-upsampling style encoder-decoder network. Each stage between the pooling layers uses dense blocks. Also, it concatenates skip connections from the encoder to the decoder.

We use these modern semantic segmentation frameworks that combine low-level and high-level features from pre-trained Tiramisu models on the Camvid [20, 21] corpus to predict our images.

B. Object Detection

Object detection is one of a computer vision technique that c. The objects can be generally identified from either images or video feeds. Object detection has been applied widely in object or people tracking, robotic maneuverability, autonomous driving, intelligent motion detection, tracking systems and so on. There are many modern deep learning architectures for object detection and one of the best deep networks is YOLO (You Only Look Once) [22].

YOLO views image detection as a regression problem, which makes its pipeline quite simple, and it is extremely fast because of this simple pipeline. It is a state-of-the-art, real-time object detection system on Pascal VOC challenge [23].

It can process a streaming video in real-time with a latency of fewer than 25 seconds. During the training process, YOLO sees the entire image and is, therefore, able to include the context in object detection. In YOLO, each bounding box is predicted by features from the entire image. Each bounding box receives 5 predictions; x, y, w, h, and *confidence*. (x, y) represents the center of the bounding box relative to the bounds of the grid cell. w and h are the predicted width and height of the whole image, and *confidence* is the probability of the event (or probability of input to fall in different classes). Commonly, object detection models also generate a confidence score for each detection. The convolutional layers of the network are responsible for extracting the features, while the fully connected layers predict the coordinates and output probabilities.

In this work, we select YOLOv3 [24] from University of Washington for object detection. It is extremely fast and accurate than all the previous YOLO versions. Moreover, it uses a few tricks to increase amount of training data and improve model accuracy, including multi-scale predictions and a better backbone classifier.

IV. Proposed Method

In this paper, we propose a new approach to evaluate transportation mobility using image recognition techniques. The benefits of a new method are not only high scalability to span evaluating areas, but also cost effective.

We start our task with selecting a video recording method. In order to coverage a wide area range, we choose a video record from a camera attached to a moving vehicle. Compared to a fixed position camera (such as a surveillance camera or CCTV), our method can collect data with more coverage areas. Moreover, as the vehicle are movable, we can freely determine the areas of interest.

After collecting the video dataset, we apply two image recognition techniques, semantic segmentation and object recognition, to extract information from the video records. The summary of observed classes and acquired outputs are shown in Table 1.

TABLE I. Applied Image Techniques and Output Details

Technique	***Observed Classes***	***Recognized Output***
Semantic Segmentation	1. Car 2. Bike 3. Person 4. Road 5. Tree 6. Other	Percentage of observed areas (in pixels)
Object Recognition	1. Car 2. Truck 3. Bike 4. Person	Numbers of observed objects

The idea of using the semantic segmentation is to evaluate positive and negative areas of the images. The positive image area are our desirable factors: (i) road and (ii) tree. On the other hand, the negative areas are undesirable factors or obstacles: (i) car, (ii) bike and (iii) people. The more positive areas affect the more transportation mobility. To indicate satisfaction and unsatisfaction by each factor, percentage of area is conducted. We calculate percentages of pixels using the equation (1).

$$percentage\ of\ pixels = \frac{pixels\ of\ the\ interested\ area}{total\ pixels\ of\ the\ image} \quad (1)$$

For the desirable factors, the more percentages of pixels that the system yields, the more satisfaction is obtained. In contrast, for the undesirable factors, the more percentages of pixels the system provides, the less satisfaction that we acquire. For the object recognition, the idea is to evaluate undesirable factors by detecting the numbers of obstacles in each class: (i) car, (ii) track, (iii) bike and (iv) people. Therefore, the greater numbers of obstacles in each class are detected, the less satisfaction it means.

In the next step, we interpret the recognition results into mobility evaluation. There are many ways to interpret the acquired results. Firstly, we can use them to compare mobility between different areas in a city; Bangkok is used as a studying case in this paper. Next, we can collect sets of recognition results from various periods of times and areas in order to compare general mobility of different cities. Finally, we can combine them to enhance a performance of the traditional QOL evaluation or reduce data survey cost. However, more researches are required.

We can apply Artificial Intelligence or Machine Learning techniques to predict QOL evaluation from the recognition results. However, it requires a lot of survey data in traditional methods and more research is also required, and hence, we left the topic as a future work.

V. Experiments

A. Dataset

The original videos used in this experiment were recorded in Sukhumvit District, Bangkok. They were taken by Iwane Laboratory [25] using the Image based Mobile Mapping System (IMS3/IMS5+).

For our experiment, we categorize the scenes into three types: large-road, medium-road and small-road scenes. The large-road scene contains a road with 6 lanes, while the medium-road and small-road scene contains 4 lanes and 2 lanes or none, respectively. We capture the video records by 2 images per second, acquiring images from each scene as shown in Table 2.

TABLE II. Selected Scenes

Road Size	***Scene***	***video length***	***no. of images***
Large	1 2 3	188 seconds 188 seconds 188 seconds	375 375 375
Medium	4 5	188 seconds 188 seconds	375 375
Small	6 7	188 seconds 122 seconds	375 244

B. Implementation and Computer Configuration

We implemented our methods on Tensorflow [26]. All of the semantic segmentation experiments were conducted on a server with an Intel® Xeon® Silver 4110 CPU @ 2.10GHz (8 Cores, 16 Threads per socket; 2 Sockets), 128 GB of memory (RAM), with GPU: Nvidia Tesla V100 32GB x 2, while all of the object detection experiments were conducted on a server with an Intel® Xeon® Processor E5-2660 v3 (25M Cache, 2.60 GHz), 32 GB of memory (RAM), an Nvidia GeForce

GTX 1070 (8 GB), an Nvidia GeForce GTX 1080 (8 GB), and an Nvidia GeForce GTX 1080 Ti (11 GB).

C. Hyperparameter Confuguration

For a semantic segmentation on CamVid [20, 21], we first remove the last three pooling layers and the last dense layer since it helps to reduce the number of parameters and computation in the deep learning network without decreasing performance. Then, we set the dilation rates of the convolution layers as 3 and 5, respectively. We also make experiments using YOLOv3 with ResNet that consists of a 3×3 convolutional layer with 64 channels, followed by 3 stages with 2 basic blocks in each stage and ends up with a global average pooling and a 10-way fully connected layer, as the backbone.

Specifically, we adopt Nesterov momentum optimizer [27] with momentum = 0.9, initial learning rate = 0.05, rate decay = 0.94 every 2 epochs, and weight decay 4e-5. Batch normalization [28] is used before each weight layer in our implementation to ease the training and make it is comparable to concatenate feature maps from different layers. To avoid overfitting, data augmentations are used as data preprocessing, including random flipping vertically, random flipping horizontally, and a random crop of 512×512 image patches.

D. Experiments and Results

We feed the dataset images into our image recognition system to extract mobility components. The examples of the large-road, medium-road and small-road scenes are shown in figures 2-7, respectively[1].

The object detection system yields excellent performance, while the semantic segmentation system gives satisfy performance. For the object detection system via the pretrain YOLOv3 network, unsurprisingly, all the classes can be detected very well. The semantic segmentation system performs very well when the image texture is clearly separated in each section, such as the images of large road. But its performance tends to decrease for medium and small road because the image dataset contains a lot of unclear edges between each section.

The summary of recognition results is shown in Table 3. In general, the trends of semantic segmentation and object detection results from the same class and scene are in the same direction. For the large road, the images contain a lot of areas and numbers of cars and roads, but few in areas of bikes and person. For the medium road, the object detection can work well. However, as there are a lot of unclear edges of shadow or dark sections, the semantic segmentation outputs higher tree areas and lower road areas than the actual. In the scene 5, the number of cars seems to be high compared to its area. This is because the system detects a lot of cars from opposite direction of the road and count into the system (as shown in fig. 2). For the small road, as they contain a lot of unidentified objects, the segmentation system recognizes some of them as bikes and person, and hence their percentage of areas seems to be too high. However, thanks to the object detection system that performs well, the actual road situation can be implied.

We can also analyze some mobility situation from the experimental results. For the large road, the main objects are cars, as the percentages of cars and roads are relatively high compared to bikes and persons. On the other hand, bikes and persons are main objects for the medium or small road. Among the scenes in the large road, scene 2 is the highest in mobility because it contains the lowest percentage of cars and contains the highest percentage of the road. For the same reason, the highest mobility scenes for the medium and small roads are scenes 4 and 7, respectively.

VI. Conclusions

The Quality of Life (QOL) is an important issue in city development. The traditional researches require questionnaire surveys that consume a lot of resources in data gathering. As the Artificial Intelligence technologies are developed rapidly, they are applied to solve many problems, including the transportation problems. In this paper, we propose an image recognition method for extracting mobility factors from videos or images. The new method consists of two techniques, semantic segmentation and object recognition. The first one recognizes percentage of positive and negative components on the observed videos or images, while the second one detects the number of obstacles. As the mobility is a significant component of QOL, our extracted factors could be added to QOL evaluation process in order to augment data or cut down the cost of data collecting.

We perform experiments and gain satisfy results in factor extraction. The acquired results can be used to indicate the actual mobility situation of the observed road and can be used in transportation planning or city development. In the view of QOL, our system yields a set of measured number that can be used as a primary information for evaluating QOL of people. However, to figure out the exact QOL value from images, further researches are required.

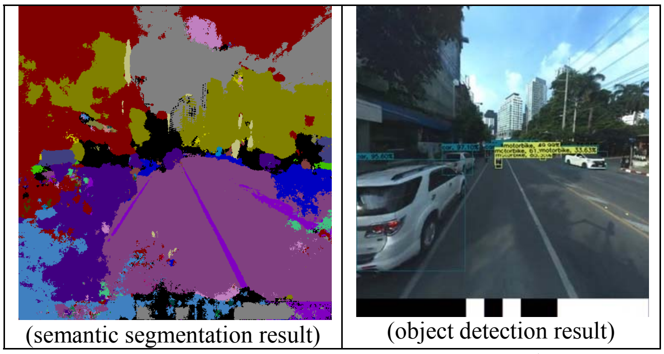
VII. Future Works

There are many possible tasks for the further works. First, although we obtain a lot of information about transportation mobility from our system, how to include them into traditional QOL evaluation are still a topic needed to be considered. Second, it is possible to predict QOL index from images using Machine Learning techniques. To achieve this task, data gathering using questionnaire survey is required. Third, in technical views, there are some objects that require more research for better classify, for example, trash bin, electric pole, footpath. Finally, in this paper, we conduct a research only in driving mode for the vehicles. Therefore, a transportation mobility in walking mode for the pedestrian can be conducted as another study.

Acknowledgment

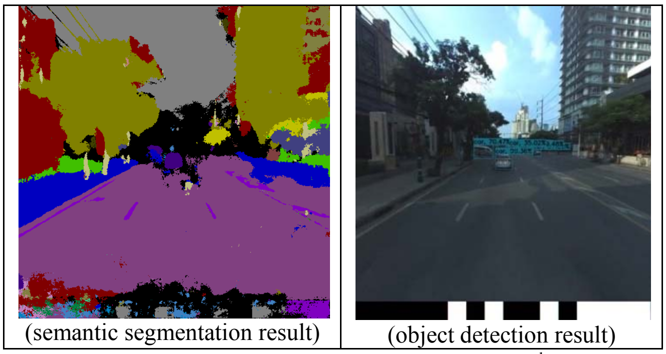
This research is supported by SATREPS Project of JST and JICA: "Smart Transport Strategy for Thailand 4.0 Realizing better quality of life and low-carbon society", by Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (C)(17K00252) and by Chubu University Grant.

[1] To see the full recognition results, please visit www.youtube.com/watch?v=W6WXxBVTKvY



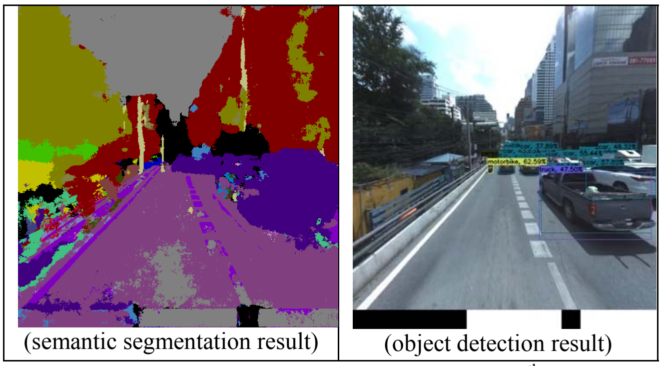
(semantic segmentation result) (object detection result)

Fig. 2 – A recognition result example from the 1st scene



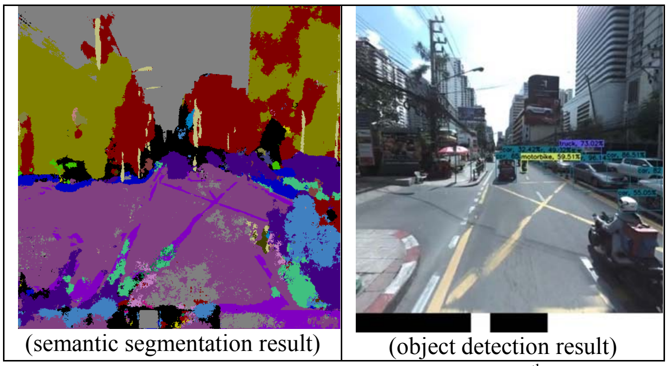
(semantic segmentation result) (object detection result)

Fig. 3 – A recognition result example from the 2rd scene



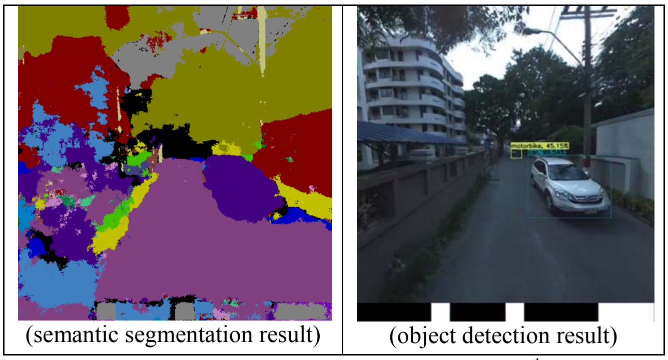
(semantic segmentation result) (object detection result)

Fig. 4 – A recognition result example from the 4th scene



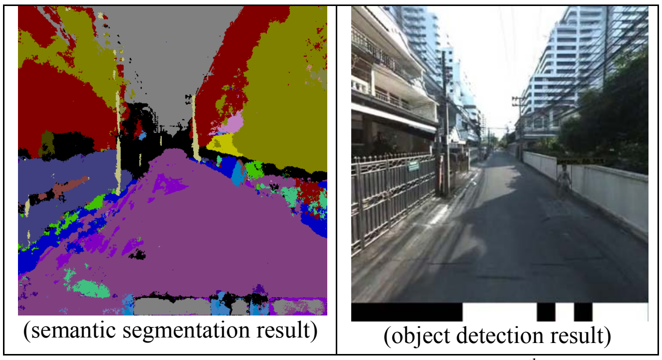
(semantic segmentation result) (object detection result)

Fig. 5 – A recognition result example from the 5th scene



(semantic segmentation result) (object detection result)

Fig. 6 – A recognition result example from the 6th scene



(semantic segmentation result) (object detection result)

Fig. 7 – A recognition result example from the 7th scene

TABLE III. IMAGE RECOGNITION RESULTS FROM ALL SCENES

Scene	Road size	Semantic Segmentation (avg. percentage of pixels)						Object Detection (avg. no. of objects)			
		Car	Bike	Person	Road	Tree	Other	Car	Truck	Bike	Person
1	Large	21.23	11.54	5.25	16.44	9.54	35.99	6.57	0.33	2.63	2.16
2	Large	18.77	9.26	8.34	22.34	7.32	33.97	6.38	0.25	0.86	0.81
3	Large	25.68	12.54	2.45	14.77	3.65	40.92	7.08	0.30	0.31	0.43
4	Medium	7.43	16.53	15.44	9.55	22.56	28.49	1.00	0.77	1.00	2.24
5	Medium	11.35	14.43	17.54	7.77	26.45	22.46	7.86	0.65	1.65	2.32
6	Small	17.54	21.77	19.45	11.56	0.23	29.43	0.78	0.04	0.09	0.40
7	Small	14.55	25.44	17.56	13.49	0.44	28.53	1.12	0.07	0.22	2.24

REFERENCES

[1] I. E. Schneider, "Quality of Life: Assessment for Transportation Performance Measures," 2013.

[2] K. Doi, M. Kii, and H. Nakanish, "An integrated evaluation method of accessibility, quality of life, and social interaction," *Environment and Planning B: Planning and Design,* vol. 35, p. 11, 2008, doi: 10.1068/b3315t.

[3] K. Nakamura, V. Wasuntarasook, F. Gu, V. Vichiensan, M. Kii, and Y. Hayashi, "Evaluation for Low-carbon Land-use Transport Development with QOL Indexes in Asian Developing Megacities: a Case Study of Bangkok," *Journal of the Eastern Asia Society for Transportation Studies,* vol. 11, 2015, doi: 10.11175/easts.11.1047.

[4] K. Nakamura, H. Morita, V. Vichiensan, T. Togawa, and Y. Hayashi, "Comparative Analysis of QOL in Station Areas between Cities at Different Development Stages, Bangkok and Nagoya," *Transportation Research Procedia,* vol. 25, pp. 3188–3202, 2017, doi: 10.1016/j.trpro.2017.05.361.

[5] F. Gu, Y. Hayashi, F. Shi, H. Zhang, and H. Kato, "Measuring and mapping the spatial distribution of the quality of life in a city: a case study in Nanjing," *International Journal of Urban Sciences,* vol. 20, no. 1, pp. 107–128, 2016, doi: 10.1080/12265934.2015.1127179.

[6] P. N. Chowdhury, T. C. Ray, and J. Uddin, "A Vehicle Detection Technique for Traffic Management using Image Processing," 2018, pp. 1-4, doi: 10.1109/IC4ME2.2018.8465599.

[7] S. Hua and D. C. Anastasiu, "Effective Vehicle Tracking Algorithm for Smart Traffic Networks," in *2019 IEEE International Conference on Service-Oriented System Engineering (SOSE)*, 2019, doi: 10.1109/SOSE.2019.00019.

[8] A. Putri, A. Achmad, and Suwadi, "Estimated Vehicle Density Based on Video Processing Using the Gaussian Mixture Model Method," *Journal of Physics: Conference Series,* vol. 1201, p. 012005, 2019, doi: 10.1088/1742-6596/1201/1/012005.

[9] E.-J. Kim, H.-C. Park, S.-W. Ham, S.-Y. Kho, and D.-K. Kim, "Extracting Vehicle Trajectories Using Unmanned Aerial Vehicles in Congested Traffic Conditions," *Journal of Advanced Transportation,* 2019, doi: 10.1155/2019/9060797.

[10] T. Osman, S. S. Psyche, J. M. S. Ferdous, and H. U. Zaman, "Intelligent Traffic Management System for Cross Section of Roads Using Computer Vision," in *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, 2017, pp. 1-7, doi: 10.1109/CCWC.2017.7868350.

[11] B.-H. Chen and S.-C. Huang, "Probabilistic neural networks based moving vehicles extraction algorithm for intelligent traffic surveillance systems," *Information Sciences,* vol. 299, 2015, doi: 10.1016/j.ins.2014.12.033.

[12] I. Ali, A. Malik, W. Ahmed, and S. A. Khan, "Real-time Vehicle Recognition and Improved Traffic Congestion Resolution," in *2015 13th International Conference on Frontiers of Information Technology (FIT)*, 2015, pp. 228-233, doi: 10.1109/FIT.2015.48.

[13] M. D. E. Munajat, R. Munir, and D. H. Widyantoro, "Traffic Congestion Model: Challenges and Opportunities," *Journal of Physics: Conference Series,* vol. 801, p. 012048, 2017, doi: 10.1088/1742-6596/801/1/012048.

[14] V.-T. Dinh, N.-D. Luu, and H.-H. Trinh, "Vehicle Classification and Detection based Coarse data for Warning Traffic Jam In VietNam," in *2016 3rd National Foundation for Science and Technology Development Conference on Information and Computer Science*, 2016, pp. 223-228.

[15] J. Trivedi, M. S. Devi, and D. Dhara, "Vehicle Counting Module Design in Small Scale for Traffic Management in Smart City," in *2018 3rd International Conference for Convergence in Technology (I2CT)*, 2018, pp. 1-6, doi: 10.1109/I2CT.2018.8529506.

[16] M. Delavarian and O. Maarouzi, "Vehicle Tracking at Intersection in Image Sequences with MultiLayer Concept," in *2017 3rd Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, 2017, pp. 131-135, doi: 10.1109/ICSPIS.2017.8311603.

[17] N. C. Mithun, T. Howlader, and S. M. M. Rahman, "Video-based tracking of vehicles using multiple time-spatial images," *Expert Systems With Applications,* vol. 62, 2016, doi: 10.1016/j.eswa.2016.06.020.

[18] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.

[19] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1175-1183, doi: 10.1109/CVPRW.2017.156.

[20] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, "Segmentation and Recognition Using Structure from Motion Point Clouds," in *ECCV (1)*, 2008, vol. 5302, pp. 44-57, doi: 10.1007/978-3-540-88682-2_5.

[21] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic Object Classes in Video: A High-Definition Ground Truth Database," *Pattern Recognition Letters,* vol. xx, pp. xx-xx, 2008.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[23] "The PASCAL VOC project." The Pattern Analysis Statistical Modelling and Computational Learning VOC project. http://host.robots.ox.ac.uk/pascal/VOC/

[24] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," 2018.

[25] "Iwane Labolatories, Ltd." http://www.iwane.com/

[26] M. Abadi *et al.*, "TensorFlow: a system for large-scale machine learning," presented at the Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation, Savannah, GA, USA, 2016.

[27] J. Lucas, R. Zemel, and R. Grosse, "Aggregated Momentum: Stability Through Passive Damping," 2018.

[28] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," presented at the Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille, France, 2015.